

Supplementary notes

1. Impact of CAD-associated SNPs on proteins.

In this analysis, we used a panel of proteins that have either a potential or an established link to CVD-related phenotypes. We therefore have specifically assessed the impact of the disease-associated SNPs on these proteins. To do so, we used all reported SNPs associated with coronary artery disease (CAD) in a recent paper by Nelson et al. ¹. The paper reported 72 known loci (73 unique SNPs) and a list of 304 suggestive loci (366 unique SNPs), with the latter partially overlapping the known ones and making a total of 422 unique CAD-associated SNPs. We then performed *cis*- and *trans*-pQTL analysis between these SNPs and 92 proteins. At FDR < 0.05 level, we identified 29 significant associations (12 unique SNPs, 14 unique proteins, **Supplementary Table 4**), including 18 *trans*-pQTLs at the *ABO* locus, a known pleiotropic locus on chromosome 9.

Next, we assessed the association between proteins and the combined genetic risk score of CAD. We used these two CAD-associated SNP lists, a short list with 72 established loci and a long list with 304 suggestive loci, and constructed a weighted genetic risk score (GRS) respectively. Plasma level of granulin (GRN) was significantly positively associated to the GRS of CAD ($P_{\text{adj}}=5.34 \times 10^{-7}$ for the GRS based on 304 loci and $P_{\text{adj}}=0.001$ for the GRS based on 72 loci), suggesting it as a biomarker for CAD development.

2. *Cis*- and *trans*-pQTLs identify driver genes in CVD

Proteins under strong *cis* genetic control are potentially interesting therapeutic targets, particularly when the *cis*-pQTL SNPs are also associated with complex traits and diseases. For example, the *cis*-pQTL SNPs we identified explain up to 73.3% of the variation in IL-6RA levels and 68.6% of the variation in IL-17RA levels. The top IL-6RA-associated SNP, rs2228145 ($r_{\text{LD1}}=0.83$; $P=3.3 \times 10^{-310}$), is a missense variant that has been associated with cardiovascular disease and many other complex traits ²¹⁻²⁴, and IL-6RA has been proposed as a potential target for CVD treatment ²⁵. We also detected strong *cis*-genetic variation for many immune-related proteins, including the signal-regulatory protein, SHPS1 (72.1%), and several members of the C-C motif chemokine

ligands: CCL15 (66.0%), CCL24 (57.3%), and CCL16 (56.6%) (**Supplementary Table 14**). Some pQTL SNPs are also associated to immune-related disease and traits, highlighting their role in complex diseases.

Trans-pQTLs can point to proteins involved in the same biological process and reveal putative drivers of these pathways. The strongest *trans*-pQTL detected in our dataset was between the missense SNP rs4760 at the *PLAUR* gene and the plasma level of TNFRSF10C ($r_{\text{LD1}} = -0.65$; $P = 3.8 \times 10^{-149}$), a link consistent with previous observations ¹¹. This effect explains 42.2% of variance in the levels of TNFRSF10C. Further, pleiotropic *trans*-pQTL effects were detected for several loci, including the *ABO*, *KLKB1*, *ST3GAL6*, *HLA*, and *FUT2* loci, and most of these observations are novel (**Supplementary Table 3**). For instance, our *trans*-pQTL analysis identified 12 *trans*-pQTLs for 4 independent SNPs in the *KLKB1* locus and 9 of these mapped to the missense allele rs3733402*G, which was associated with higher levels of PCSK9, NT-pro-BNP, EPHB4, OPN, U-PAR and MEPE and lower levels of CDH5, LTBR and TFPI (**Supplementary Fig. 4A**). These proteins showed modest to low inter-correlation, suggesting a pleiotropic effect of the *KLKB1* locus (**Supplementary Fig. 4B**). Moreover, these proteins are physically located in GWAS loci and genetic variants in all of these proteins are linked to CVD and its related risk factors, including heart rate, blood pressure, blood lipid level and blood cell counts (**Supplementary Fig. 4C**). Previous GWAS analysis and pQTL studies have linked rs3733402-G to high risk of CVD²⁶, to higher circulating levels of two established protein markers for heart failure (BNP and pro-BNP)²⁷, to higher levels of serum free insulin-like growth factor 1 (IGF-1)²⁷, and to lower levels of several serum metabolites ²⁸. The *KLKB1* gene encodes plasma kallikrein, a proteolytic enzyme known to cleave high-molecular-weight kininogen to bradykinin and prorenin to renin ^{29,30}. Kallikrein also affects other vasoactive hormones such as endothelin-1 and midregional proadrenomedullin ³¹. Bradykinin and renin are key enzymes in vasomotion and in blood pressure, water, and salt homeostasis ³¹. The SNP rs3733402 causes an amino acid substitution, Asn124Ser, in the heavy chain apple 2 domain of kallikrein ^{32,33}. This substitution reduces kallikrein activity by affecting its binding to cofactors and substrates ^{34–36}. Therefore, the pleiotropic *trans*-pQTLs we observed in the *KLKB1* SNP are likely mediated via changes in kallikrein activity, a mechanism which has been suggested before ^{31,37,38}.

3. Overlap of pQTLs with eQTLs

The low overlap between pQTLs and eQTLs we observe is consistent with several other studies that reported a low correlation between protein and transcript abundance in blood 2,3 . This discordance may be due to factors such as differences in tissue of origin or post-transcriptional modifications when variants affect protein levels rather than transcript abundance, or they may be due to technical issues such as differential protein binding. Given our large sample of proteomics and RNA-seq expression data from the same individuals (n=1,180), we were able to examine expression–protein level correlation and pQTL–eQTL overlap in our data.

We first checked the correlation of plasma protein concentrations with the expression levels of their coding genes determined by RNA-seq profiling in the same samples in whole blood 4. We calculated Spearman's correlation between pairs of proteins and the expression of the corresponding coding gene. At FDR 0.05 level, significant correlation was detected for a total of 26 gene-protein pairs, including 22 positive correlations (**Supplementary Table 6**). Most protein-gene correlations were modest to small, with the top correlations observed for CHIL3LI ($r=0.35$, $P=9.55 \times 10^{-36}$), CSTB ($r=0.24$, $P=2.54 \times 10^{-17}$) and RETN ($r=0.20$, $P=6.78 \times 10^{-12}$). The discordance highlights the added value of the plasma proteomics data as well as the other *omics* data in blood.

A similar discordance was also observed at genetic effect level. We checked the overlap of pQTLs and eQTLs for the same samples and found that 58 of 129 (45%) *cis*-pQTLs have a corresponding eQTL that is significant at FDR 0.05, but only 38 of these 58 (66%) had the same effect direction (**Supplementary Fig. 5**). In line with the high correlation between protein levels and transcript abundance, the most concordant *cis*-pQTL and *cis*-eQTL pair was detected for CHI3L1: rs4950928-G *cis*-affected the gene expression (*cis*-eQTL Z score=-0.21) and protein concentration in plasma (*cis*-pQTL Z=-18.8) (**Supplementary Table 2 & Supplementary Fig. 5**). To take different tissue types into account, we further included the eQTL data of different tissue types from the GTEx project (version 7) 5. This added 25 more overlapping eQTLs and resulted in 63 out of 83 (76%) pQTLs with a corresponding eQTL overlapping genetic effect on protein and gene expression levels with the same allelic direction. Interestingly, for all 85 *trans*-pQTLs detected in our study, we did not find concordance with *trans*-effects at transcript level either in blood or in other tissues in GTEx. At the same time, our *trans*-pQTLs showed 88% replication rate with published *trans*-pQTLs. Overall, our data

indicate that the power of *trans*-effects is higher at protein level than at gene expression level.

4. Overlap of pQTLs with mbQTLs.

To gain more insight into the relationship between genetic and microbial associations, we systematically compared the overlap of pQTLs and mbQTLs previously reported by Bonder et al.⁶

First, we investigated the overlap between all genome-wide significant mbQTLs and pQTLs. We focused on all protein-associated microbial factors and extracted 26 genome-wide significant mbQTLs for 9 taxa and 14 bacterial pathways. None of the mbQTL SNPs show linkage disequilibrium with the identified pQTL SNPs at $r^2 > 0.8$ level. This is in line with our observation that genetic and microbial associations are largely independent. However, this may partly be explained by insufficient power, as a large number of genetic variants may have very small effects on the gut microbiome that remain to be discovered.

Second, we focused on 31 proteins that were affected by both genetics and microbiome features and investigated the impact of their pQTLs SNPs on the microbiome features. For these 31 proteins, we extracted their associations with 89 pQTL-SNPs (50 *cis*- and 39 *trans*-pQTLs), 63 unique bacterial taxonomies (**Supplementary Table 9**) and 175 unique bacterial pathways (**Supplementary Table 10**). We further compared the associations between these 89 pQTL SNPs and the bacterial taxonomies and pathways and reported a total of 18 suggestive associations for 7 pQTL SNPs at a nominal P-value 0.05 level (**Supplementary Table 11**). These include three *trans*-pQTL of Ep-CAM at the FUT2 locus affecting six taxonomies and three pathways and two *cis*-pQTL SNPs of PON3 affecting one taxonomy and five pathways. Moreover, two microbial associations were found for a *cis*-pQTL SNP of PAI and one association for *cis*-pQTL of CHI3L1. Interestingly, these proteins are among the top proteins associated with the gut microbiome. The strongest association was observed between the *FUT2* locus and *Blautia*, an association which has been replicated in many other studies (see the discussion in the main text). Other associations between pQTLs and microbial factors need to be further validated.

Third, we further investigated to what extent the associations between protein and microbiome could be influenced by genetic effects. We recalculated microbiome-protein associations after regressing out all genome-wide significant mbQTLs and pQTLs, i.e. 26

mbQTLs and 224 pQTLs (**Supplementary Tables 12 & 13**). The association strengths before and after correction for genetic effects were very comparable (**Supplementary Fig. 8**). The top difference was observed for the negative association between Ep-CAM and the superpathway of L-phenylalanine and L-tyrosine biosynthesis (PWY-3481). After regressing out the pQTLs and mbQTLs, the correlation coefficient of association increased from -0.1 ($P=8.9 \times 10^{-4}$) to -0.15 ($P=4.1 \times 10^{-6}$).

Altogether, the impact of pQTLs on microbial factors and on protein-microbial associations was mostly negligible.

5. Assessment of the independent and additive effects of genetic and microbial factors on plasma proteins

Next, we compared the total explained variance in proteins using two different models. Model 1 assumes that the genetic and microbial factors were completely independent, i.e. the total explained variation could be estimated by summing the variation explained by genetic factors ($V_g = V_{cis} + V_{trans}$) and microbial factors (V_m) using two separate models. Model 2 assumes the genetic and microbial factors were not fully independent, i.e. the total explained variation (V_t) was estimated by a combined model including both genetic and microbial factors. The difference between the two models would indicate to what extent the genetic effects and microbial effects were independent or confounding. We observed a very high concordance between the two models ($r=0.99$, $P=0$), which suggests genetic and microbial effects are independent for most proteins. The greatest difference between our two models was for intestinal epithelial protein Ep-CAM, where 1% of the total variation in Ep-CAM may be due to a confounding effect of genetic and microbial factors.

To further study the dependency between genetic and microbiome effects on protein levels, we searched for interactions between SNP genotypes and microbial factors.

We detected 21 nominally significant ($P < 0.05$) genome-microbiome interactions for 8 proteins (**Supplementary Table 15**): one interaction each for CNTN1, CHI3L1, CTSZ, PON3, IL6-RA, PSP-D, TLT-2 and 14 significant interactions for Ep-CAM. For these proteins, the associations between microbiome feature and protein level differ among genotypes. For example, the vitamin B1 pathway (PWY-7357: thiamin formation) is only significantly correlated with PON3 in rs10953142 C/C genotype individuals. One study has reported a decrease in HDL cholesterol after six months of thiamin supplementation ⁷, which aligns with our observation of positive association between

PON3 and HDL (see main text). Ep-CAM is negatively associated with many beneficial bacteria and pathways, including *Bifidobacterium*, the Peptostreptococcaceae family and the Coenzyme B synthesis pathway. However, this association seems to have a larger effect in homozygous individuals for the SNPs in the FUT2 locus.

Through explained variance estimation and interaction analysis, we found that genetics and microbiome affect protein levels mainly in an additive manner. However, our data also provides evidence for functional interactions between host and microbiome that affect plasma protein concentrations.

References

1. Nelson, C. P. *et al.* Association analyses based on false discovery rate implicate new loci for coronary artery disease. *Nat. Genet.* **49**, 1385–1391 (2017).
2. Chick, J. M. *et al.* Defining the consequences of genetic variation on a proteome-wide scale. *Nature* **534**, 500–505 (2016).
3. Sun, W. *et al.* Common Genetic Polymorphisms Influence Blood Biomarker Measurements in COPD. *PLoS Genet.* **12**, e1006011 (2016).
4. Zhernakova, D. V. *et al.* Identification of context-dependent expression quantitative trait loci in whole blood. *Nat. Genet.* **49**, 139–145 (2016).
5. Aguet, F. *et al.* Genetic effects on gene expression across human tissues. *Nature* **550**, 204–213 (2017).
6. Bonder, M. J. *et al.* The effect of host genetics on the gut microbiome. *Nat Genet* **48**, 1407–1412 (2016).
7. Al-Attas, O. *et al.* Metabolic benefits of six-month thiamine supplementation in patients with and without diabetes mellitus type 2. *Clin. Med. Insights Endocrinol. Diabetes* **7**, (2014).